

知识图谱约束深度网络的高分辨率遥感影像场景分类

李彦胜, 吴敏郎, 张永军

武汉大学遥感信息工程学院, 湖北 武汉 430079

收稿日期: 2023-04-23; 修回日期: 2024-01-15 中图分类号: P237 文献标识码: A

文章编号: 1001-1595(2024)04-0677-12

基金项目: 国家自然科学基金(42030102; 42371321)

第一作者简介: 李彦胜(1987—), 男, 博士, 教授, 研究方向为遥感时空知识图谱、多模态遥感基础大模型、遥感大数据智能解译。

E-mail: yansheng.li@whu.edu.cn

摘要: 得益于深度网络理论与方法的快速发展, 深度网络逐渐成为遥感影像场景分类任务的主流技术。然而, 现有基于深度网络的遥感影像场景分类方法高度依赖大量人工标记的训练样本, 且无法有效融合利用遥感领域丰富的先验知识。为了提升领域知识利用率同时降低标记样本依赖, 本文提出了一种知识图谱引导深度网络学习的高分辨率遥感影像场景分类方法。首先, 构建了一个包括领域内多种来源知识的土地覆盖概念知识图谱来更灵活便捷地应用领域先验知识。然后, 通过知识图谱表示学习方法将土地覆盖概念知识图谱中的遥感场景语义类别表达为语义向量, 形成遥感场景类别语义基准。在知识引导学习阶段, 通过施加场景类别语义向量与深度网络浅层视觉特征向量的跨模态对齐约束引导深度网络的浅层部分更有效地学习不同类别遥感影像场景的共享特征, 在深度网络深层部分则仍然通过场景类别标签引导学习不同遥感场景的判别特征。在测试阶段, 完成优化的深度网络模型可以在不依赖任何先验知识的情况下完成高精度遥感影像场景分类。在目前公开的最大的遥感影像场景分类数据集上的试验结果表明, 本文提出的知识引导学习方法相比现有方法在 10%、30%、50% 等不同训练样本比率下均可以获得最佳分类性能。在 10% 这一比率条件下, 本文提出的知识引导学习方法相比基线深度网络在总体精度指标(OA)上能够得到 5.11% 的提升。

关键词: 遥感影像场景分类; 土地覆盖概念知识图谱; 知识图谱表示学习; 知识图谱约束深度网络

由于高分辨率遥感影像存在典型的“同物异谱”“同谱异物”现象, 传统的基于像素或对象的遥感影像地物分类方法受到很大的局限性, 难以有效满足遥感大数据时代下对高分辨率遥感影像解译能力的需要与应用的需求。基于这一现实, 遥感影像场景分类任务越来越多地受到遥感领域研究者的关注^[1-3]。遥感影像场景分类旨在挖掘遥感影像上基本视觉单元间的联系, 从而对影像整体所属的遥感场景语义类别进行预测, 有效提高高分辨率遥感影像解译工作的稳定性与可靠性。遥感影像场

景分类在灾害监测^[4-5]、土地覆盖分类调查^[6-7]等任务中起着至关重要的作用。

现有的遥感影像场景分类方法主要包括基于人工特征的传统分类方法^[8-9]与基于数据驱动的深度学习方法^[10-11]。随着 AID30^[12]、NWPU-RESISC45^[13]、PatternNet^[14] 等高分辨率遥感影像数据集的构建和公开, 数据学习模式带来的高层视觉特征抽取能力使得基于数据驱动的深度学习方法在遥感影像场景分类任务上取得了优异的性能表现。现有基于深度学习的遥感影像场景分类方法主要分为两

引文格式: 李彦胜, 吴敏郎, 张永军. 知识图谱约束深度网络的高分辨率遥感影像场景分类[J]. 测绘学报, 2024, 53(4): 677-688. DOI: 10.11947/j. AGCS. 2024. 20230125.
LI Yansheng, WU Minlang, ZHANG Yongjun. Knowledge graph-guided deep network for high-resolution remote sensing image scene classification[J]. Acta Geodaetica et Cartographica Sinica, 2024, 53(4): 677-688. DOI: 10.11947/j. AGCS. 2024. 20230125.

类:第1类是基于传统卷积网络的模型^[15-17];第2类是基于多头注意力机制的视觉 Transformer 模型^[18-19]。但是,上述基于深度学习的方法高度依赖在充足的人工标记训练样本上对模型进行训练,在面对不充分的训练样本时很容易出现对训练样本过拟合的问题,限制了其在遥感影像场景分类任务中的表现^[20]。

为了克服深度学习模型高度依赖大量训练样本的局限性,通过先验知识来引导深度网络进行学习是一个得到国内外权威专家认可,拥有极大优势与研究潜力的思路。在学术领域,有较多引入先验知识解决遥感领域问题的学术成果,并取得了显著的成效。具体来说,文献[21]通过遥感场景视觉特征与语义表示之间的稳健跨模态匹配,解决零样本条件下的遥感影像分类问题。文献[22]通过引入本体推理模块,通过先验知识本体推理引导深度学习网络得到更好的遥感影像语义分割结果。文献[23]提出了一种用于土地覆盖分类任务的领域知识引导深度协同融合网络,通过领域知识改善深度网络对少数类的分类性能。文献[24]通过建立一个类层次树来引导深度网络模型充分地利用不同地表覆盖类别的语义信息,提高深度网络模型的高光谱遥感影像分类性能。文献[25]提出了知识引导的土地格局绘制框架,通过领域内先验知识引导城市土地利用制图框架生成更准确的城市土地利用图。但是,在高分辨率遥感影像场景分类方面,如何对领域先验知识进行有效管理并生成场景类别语义基准,并用于引导深度网络学习,仍是一个亟待研究的问题。

为了更有效地管理遥感领域先验知识,知识图谱近年来受到领域内学者的广泛关注。知识图谱是一种通过节点与边来对客观世界中的概念及概念间相关关系进行描述的语义网络^[26]。知识图谱能够以一种更加贴近人类视角的形式来建模计算机世界的信息与知识,被广泛应用于语义搜索、智能问答、预测决策等重要的人工智能相关任务中^[27]。目前有许多开源领域无关知识图谱,例如 Freebase^[28]、DBpedia^[29]、CN-Probase^[30]等。然而,领域无关知识图谱更侧重于对人类知识覆盖的广度,面对特定领域的知识应用任务时,领域无关知识图谱的领域专业性往往不能满足领域知识应用任务的需要。为了解决这一问题,针对遥感领域中具体的任务,一些研究者设想并构建起了与遥感、地学、地理等领域相关的领域知识图谱^[31-37],并应用到具体领域知识应用任务中。但是,上述领域知

识图谱尚未针对遥感场景类别进行概念级建模,因此无法对遥感场景类别进行准确描述。对于遥感场景语义基准,现有研究普遍采用自然语言处理模型来进行表达生成,例如 Word2Vec 模型^[38]、Bert 模型^[39]等。由于自然语言处理模型往往在领域无关的语料上进行学习,其得到的语义基准往往忽视了领域特殊性,无法准确建模遥感场景特点。因此,需要基于土地覆盖概念知识图谱得到更佳的语义表达,现有研究主要通过知识图谱表示学习方法来达到这一目的。现有知识图谱表示学习方法主要包含:①基于距离的翻译模型,如 TransE^[40]、TransH^[41]、TransR^[42]、RotatE^[43]和 HAKE^[44]等;②基于实体语义相似度的语义匹配模型,如 RESCAL^[45]、DistMult^[46]、CompLex^[47]和 TuckER^[48]等。上述通用领域知识图谱表示学习方法在遥感领域的适用性尚需探索研究。

基于上述考虑,本文提出一种知识图谱引导深度网络学习的高分辨率遥感影像场景分类方法。通过土地覆盖分类概念知识图谱来管理领域先验知识,并通过知识图谱表示学习方法生成遥感场景语义基准。进一步,通过遥感场景语义基准引导深度网络学习实现遥感场景分类。在目前公开最大遥感影像场景分类数据集上的试验结果表明,本文提出的知识图谱引导深度网络学习的高分辨率遥感影像场景分类方法相比现有方法在 10%、30%、50%等不同训练样本比率下均可以获得最佳分类性能。

1 知识图谱引导深度网络的高分辨率遥感影像场景分类方法

为了更加清晰地描述本文提出的知识图谱引导深度网络学习的高分辨率遥感影像场景分类方法,首先对遥感影像场景分类任务进行符号定义。给定一个遥感影像数据集 $\mathbf{H} \in \mathbf{R}^{H \times W \times C}$,其中 H 、 W 和 C 代表每张遥感影像的高度、宽度和通道数。遥感影像数据集 \mathbf{H} 包含了总共有 N 张带有场景语义类别标签的影像 $\{x_1, x_2, \dots, x_N\} \in \mathbf{R}^{H \times W \times C}$,以及影像对应的影像场景语义类别标签 $\mathbf{Y} = \{y_1, y_2, \dots, y_N\} \in \mathbf{R}^K$,其中 K 为遥感场景语义类别数。遥感影像场景分类模型的任务就是在给定遥感影像的情况下,为每张影像预测正确的遥感场景类别标签。

本文提出的知识图谱引导深度网络学习的高分辨率遥感影像场景分类方法如图 1 所示,主要包括土地覆盖概念语义表达和知识图谱引导学习两个

部分。土地覆盖概念语义表达过程主要包含土地覆盖概念知识图谱 G 构建及基于知识图谱表示学习得到遥感场景类别语义基准 $A = \{a_1, a_2, \dots, a_K\}$, 具体

方法在 1.1 节描述。知识图谱引导学习过程利用遥感场景类别语义基准 A 引导深度网络进行遥感影像场景视觉特征学习, 具体方法在 1.2 节描述。

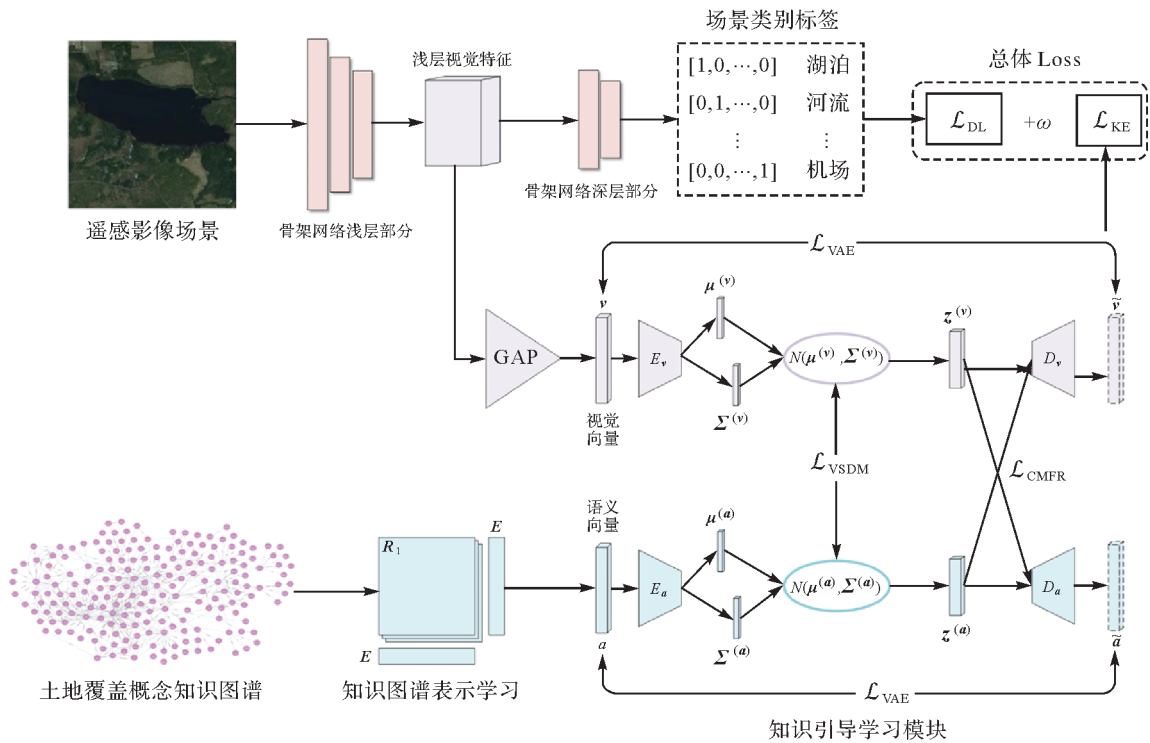


图 1 算法整体框架

Fig. 1 Framework of the proposed algorithm

遥感场景类别语义基准可能使得深度网络模型难以学习到语义相似性较高的遥感场景类别间的判别特征, 导致遥感场景分类模型的最终精度表现下降。为了解决这一问题, 在训练过程中, 通过知识引导学习模块对深度网络浅层部分施加视觉特征与语义向量 A 进行跨模态对齐约束, 引导深度网络浅层部分更有效地学习遥感场景类别间的共享特征, 该部分通过知识嵌入损失 \mathcal{L}_{KE} 来引导学习。在深度网络的深层部分, 仍然通过遥感场景类别标签 Y 进行引导, 保持深度网络学习遥感场景类别间判别特征的能力, 该部分通过数据学习损失 \mathcal{L}_{DL} 来引导学习。数据学习损失采用常规的交叉熵损失函数。本文提出的知识图谱引导深度网络学习方法的总体损失函数如式(1)所示

$$\mathcal{L} = \mathcal{L}_{DL} + \omega \mathcal{L}_{KE} \quad (1)$$

式中, ω 为损失平衡权重, 为了平衡上述两个不同的部分引导深度网络模型学习的效果。

1.1 土地覆盖概念知识图谱构建及土地覆盖概念语义表达

为了实现遥感场景先验知识引导深度网络学

习, 首先需要构建土地覆盖概念知识图谱, 对遥感场景先验知识进行更全面、更灵活、更便捷的管理与利用。在土地覆盖概念知识图谱的基础上, 通过基于知识图谱表示学习的土地覆盖概念语义表达方法, 将遥感场景先验知识映射为遥感场景类别语义向量, 形成遥感场景语义基准, 服务于后续的遥感影像场景分类任务。

1.1.1 土地覆盖概念知识图谱构建

为了管理与利用遥感场景先验知识, 本文首先构建土地覆盖概念知识图谱。在构建遥感土地覆盖概念知识图谱时, 充分考虑了领域标准、专家知识、通用知识等多种知识类型。领域标准主要包括国土三调分类体系, 该体系按二级分类结构将土地覆盖分类任务中的遥感场景类别进行了层次化分类, 并对每一种类别进行了严格的文字定义。根据这一领域标准, 土地覆盖概念知识图谱能够组织起遥感场景类别的语义层次信息, 并根据类别定义得到较为准确的遥感场景类别属性信息, 确保本体结构上的领域专业性。专家知识主要包括领域专家在遥感领域实践任务中整理出来的领域知识, 包含

遥感场景的常见内部结构和遥感场景间的空间关系等知识。通过专家知识构建土地覆盖概念知识图谱,能够使知识图谱以专业视角来对土地覆盖概念进行描述,同时以面向任务的形式对遥感场景先验知识进行组织。本文通过开发 Web 软件实现多人在线协同编辑的方法构建土地覆盖概念知识图谱,避免从单一领域任务视角搜集专家知识的局限性,确保土地覆盖分类知识图谱的泛化能力。经过领域标准与专家知识构建土地覆盖概念知识图谱后,该知识图谱已具有较强的知识专业性。为了进一步增强土地覆盖概念知识图谱对遥感场景先验知识表达的完备性,引入通用知识^[49]来填补与遥感场景相关的人类常识知识。具体地,根据开源概念知识图谱 ConceptNet^[50]检索得到与遥感场景类别相关的常识三元组,加入构建起的土地覆盖概念知

识图谱中,补充人类常识场景下对遥感场景概念及相关关系的事实描述。

基于上述构建原则,本文构建的土地覆盖概念知识图谱包含实体 8521 个、三元组事实 13 892 条。土地覆盖概念知识图谱示例如图 2 所示,以“跑道(runway)”为例,根据领域标准可以得到事实三元组“跑道(runway)是…的一部分(is part of)→机场用地(airport)”;根据专家知识可以得到事实三元组“跑道(runway)由…组成(is component)→水泥(cement)”;根据通用知识可以得到事实三元组“跑道(runway)是(is A)→平台(platform)”。该示例表现出通过领域标准、专家知识和通用知识,土地覆盖概念知识图谱能够从多个角度得到对遥感场景类别的描述。

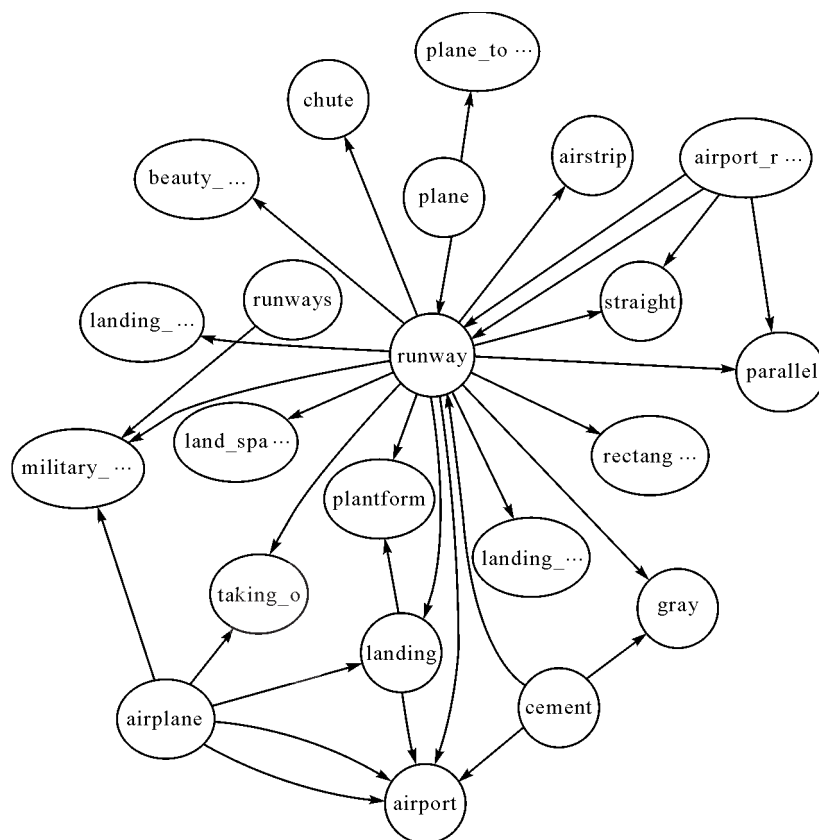


图 2 土地覆盖概念知识图谱

Fig. 2 The land cover concept knowledge graph

1.1.2 基于知识图谱表示学习的土地覆盖概念语义表达

相比 TransE^[40]、HAKE^[44]、CompEx^[47] 等知识图谱表示学习方法,语义匹配模型 DistMult^[46] 具有能够有效捕捉实体间的语义相似度和有效建模多跳关系连接的实体语义信息等优势。基于这一

考虑,本文采取 DistMult 模型来进行土地覆盖概念语义表达(图 3),并获得遥感场景语义类别表示向量。构建起的遥感场景知识图谱 G 可以被表示为事实三元组的集合 $G = \{ \langle h, r, t \rangle \mid h, t \in E, r \in R' \}$, 其中 E, R' 分别代表遥感场景知识图谱的实体集合与关系集合, h, r, t 分别代表三元组的头实体、关系

与尾实体。DistMult 模型采用一个基于双线性模型的三元组得分函数来衡量同种关系连接的头尾实体对之间的语义相似度

$$S(h, r, t) = \mathbf{y}_h^\top \mathbf{M}_r \mathbf{y}_t \quad (2)$$

式中, \mathbf{M}_r 代表关系 r 所对应的变换矩阵; \mathbf{y}_h 、 \mathbf{y}_t 分别代表头尾实体的语义表示向量。对于一个遥感场景知识图谱中存在的三元组, $S(h, r, t)$ 的得分应当接近于 1, 而对于不存在的三元组, $S(h, r, t)$ 应当接近于 0。考虑到土地覆盖概念知识图谱具有较多实体与关系, 如果不加限制, 关系变换矩阵 \mathbf{M}_r 的参

数量将会相当巨大, 不利于知识图谱表示模型的训练。因此将关系变换矩阵 \mathbf{M}_r 限制为对角矩阵, 减少模型参数量, 便于进行土地覆盖概念语义表达。

通过上述基于知识图谱表示学习的土地覆盖概念语义表达方法学习得到土地覆盖概念知识图谱中的实体语义表示向量为 $\mathbf{Y}_e = \{\mathbf{y}_{e_1}, \mathbf{y}_{e_2}, \dots, \mathbf{y}_{e_N}\}$, 其中 N 为遥感场景知识图谱中的实体总数。在 \mathbf{Y}_e 中筛选出 K 个遥感场景语义类别对应的实体语义向量, 得到遥感场景语义类别向量集合 $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K\}$ 。

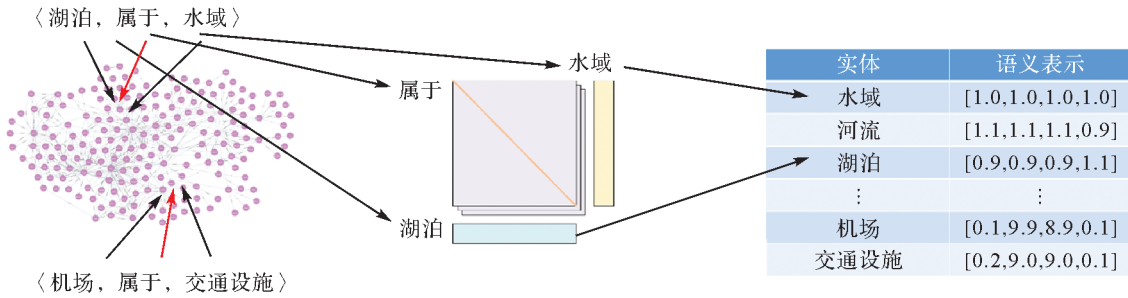


图 3 基于知识图谱表示学习的土地覆盖概念语义表达

Fig. 3 Semantic representation of land cover concepts based on knowledge graph representation learning

1.2 土地覆盖概念知识图谱引导深度网络学习

得到遥感场景语义类别向量集合 \mathbf{A} 后, 需要将遥感场景语义类别向量用于引导深度网络进行遥感影像场景分类。考虑到语义相关性较高的遥感场景往往共享更多的视觉基元, 本文通过语义引导学习模块施加遥感场景语义向量与深度网络浅层特征的跨模态对齐约束, 引导深度网络浅层部分学习到更有效的共享视觉特征。

考虑到变分自编码器 (VAE) 具有能够有效重构特征隐层空间分布和重构出的隐层空间分布具有连续性等特性与优势, 本文首次提出将 VAE 模型用于完成深度网络浅层视觉特征与语义特征进行特征重构与跨模态对齐, 从而实现通过遥感场景类别语义向量来引导深度网络浅层部分进行共享特征学习。具体实现时, 本文将骨架网络浅层部分得到的视觉特征块引出, 通过一个全局平均池化层将视觉特征块变换为视觉特征向量后, 输入知识引导学习模块进行视觉特征向量与语义类别向量的跨模态对齐学习。

知识引导学习模块中的两个 VAE 模型需要分别对输入的特征进行重构, 并提取出其特征所对应的均值向量 $\boldsymbol{\mu}$ 与方差向量 $\boldsymbol{\Sigma}$ 。为了使得 VAE 模型能够正确地重构特征, 采取基础 VAE 损失来约束 VAE 进行训练

$$\begin{aligned} \mathcal{L}_{\text{VAE}} = & \sum_i^V \mathbb{E}_{q_{\phi}^v(\mathbf{z}|\mathbf{x}^{(i)})} [\ln p_{\phi}^v(\mathbf{x}^{(i)} | \mathbf{z})] - \\ & D_{\text{KL}}(q_{\phi}^v(\mathbf{z} | \mathbf{x}^{(i)}) \| p_{\phi}^v(\mathbf{z})) + \\ & \sum_j^A \mathbb{E}_{q_{\phi}^a(\mathbf{z}|\mathbf{y}^{(j)})} [\ln p_{\phi}^a(\mathbf{y}^{(j)} | \mathbf{z})] - \\ & D_{\text{KL}}(q_{\phi}^a(\mathbf{z} | \mathbf{y}^{(j)}) \| p_{\phi}^a(\mathbf{z})) \end{aligned} \quad (3)$$

式中, \mathbf{x} 与 \mathbf{y} 分别代表原始视觉与语义向量输入; q_{ϕ}^v 与 q_{ϕ}^a 分别代表该特征对应的视觉与语义 VAE 编码器; p_{ϕ}^v 与 p_{ϕ}^a 分别代表对应的视觉与语义 VAE 解码器; D_{KL} 代表 KL 散度计算函数。通过该损失, 约束视觉特征 VAE 与语义特征 VAE 学习到对应特征向量的重构方式, 能够获得具有较强表达力的特征隐层空间向量, 提高了根据特征隐层分布对原始向量的恢复能力。

在通过基础 VAE 损失分别学习视觉特征向量与语义向量隐层空间分布的同时, 需要对能够对应的视觉特征向量与语义类别向量进行跨模态重构。跨模态损失约束如下

$$\begin{aligned} \mathcal{L}_{\text{CMFR}} = & \sum_i^N |\mathbf{x}^{(i)} - p_{\phi}^v(q_{\phi}^a(\mathbf{y}^{(i)}))| + \\ & |\mathbf{y}^{(i)} - p_{\phi}^a(q_{\phi}^v(\mathbf{x}^{(i)}))| \end{aligned} \quad (4)$$

式中, N 代表训练样本总数; $\mathbf{x}^{(i)}$ 与 $\mathbf{y}^{(i)}$ 分别代表同一训练样本所对应的视觉特征与对应语义类别向量。通过这种交叉编码重构, 将视觉特征编码后还

原至语义向量空间,将语义类别向量编码后还原至视觉特征空间,并进行对齐,能够进一步增强知识引导学习模块的跨模态对齐能力。

进一步地,模型还需要学习对齐视觉特征与对应语义向量在隐层空间的分布。对于视觉 VAE 部分与语义 VAE 部分而言,其隐层空间的分布表示为原始特征向量在隐层空间的均值与方差 $\mu^{(v)}$ 、 $\Sigma^{(v)}$ 与 $\mu^{(a)}$ 、 $\Sigma^{(a)}$,因此,通过式(5)来匹配上述视觉特征向量与语义类别向量

$$\mathcal{L}_{\text{VSDM}} = \sum_i^N \sqrt{\|\mu_i^{(v)} - \mu_i^{(a)}\|_2^2 + \|\sqrt{\Sigma_i^{(v)}} - \sqrt{\Sigma_i^{(a)}}\|_F^2} \quad (5)$$

式中, N 代表训练样本总数; $\mu_i^{(v)}$ 、 $\Sigma_i^{(v)}$ 分别代表样本视觉特征向量对应隐层空间的均值与方差; $\mu_i^{(a)}$ 、 $\Sigma_i^{(a)}$ 分别代表样本对应语义类别向量对应隐层空间的均值与方差; $\|\cdot\|_2$ 与 $\|\cdot\|_F$ 分别代表 2 范数与 F 范数。

知识引导学习模块通过上述 3 个损失,在隐层空间学习并重构视觉特征向量与语义类别向量的同时,将其进行跨模态对齐,从而通过语义类别向量引导骨架网络浅层部分学习共享视觉特征。知识引导学习模块的总体损失函数如下

$$\mathcal{L}_{\text{KE}} = \mathcal{L}_{\text{VAE}} + \alpha \mathcal{L}_{\text{CMFR}} + \beta \mathcal{L}_{\text{VSDM}} \quad (6)$$

式中, α 与 β 为权重平衡参数。经过试验验证分析,两个参数敏感性较低,本文将分别固定为 1 和 10。

2 试验及结果分析

2.1 数据集以及试验设置

为了验证本文方法的有效性,本文试验部分采用了当前公开最大的遥感影像场景分类数据集^[21],如图 4 所示。该遥感影像场景分类数据集中包含 70 类遥感影像场景,每类包括 800 张影像,每张遥感影像的像素大小为 256×256 ,是目前规模最大的遥感影像场景分类数据集,能够充分衡量本文提出的遥感影像场景分类方法的分类能力。在进行遥感影像场景分类时,由于骨架深度网络要求的输入为 224×224 ,因此先将遥感影像缩放为 224×224 后输入遥感影像场景分类模型。训练时,遥感影像场景分类模型的 batch size 设置为 16。本文的计算硬件资源为 Inter I7 3.2 GHz CPU、32 GB RAM 和 GTX2070 显卡。

本文提出的知识图谱引导深度网络学习的遥感影像场景分类方法的骨架深度网络采用预训练

的 efficientnet-b3^[51]。为了验证本文提出知识引导学习模块能够降低深度网络对于训练样本的依赖,分别设置了 10%、30%、50% 比例样本作为训练集的试验,验证集与测试集则将剩余样本均等划分形成。为了评估方法的性能,采用总体精度(OA)作为评价指标。

为了验证本文提出土地覆盖概念知识图谱的有效性,对比自然语言处理模型 Word2Vec^[38] 与 Bert^[52] 得到的遥感场景语义类别向量引导深度网络学习的遥感影像场景分类结果。在对比试验中,将本文提出的知识图谱引导深度学习的遥感影像场景分类方法与现有遥感影像场景分类方法进行对比,验证本文方法的先进性。此外,对比试验中加入了 PCNet^[15] (ResNet-101) 与 VGG16^[53] 网络作为骨架深度网络的试验,验证知识引导学习模块的应用灵活性。

2.2 关键参数分析

本文提出的知识图谱引导深度网络学习的遥感影像场景分类模型主要包含两个参数:①平衡数据学习损失 \mathcal{L}_{DL} 与知识嵌入损失 \mathcal{L}_{KE} 的权重 ω ;②知识图谱表示学习得到的语义类别向量维度。因此分别对这两个参数在 10%、30% 与 50% 条件下进行分析。

对损失平衡权重 ω 进行分析时,将知识图谱表示学习向量固定为 3000 维, ω 的取值范围为 $\omega = \{0.00001, 0.0001, 0.001, 0.01, 0.1\}$,参数分析的结果如图 5(a) 所示。可以看到,当 ω 取值较大时模型精度存在明显下降,在综合考虑了 3 种比例训练样本下的表现后,本文将 ω 设定为 0.0001。

对知识图谱表示学习向量维度进行分析时,按照损失平衡权重 ω 的分析结果,将 ω 设置为 0.0001 后,在 3 种不同的训练集划分下对不同维度表示学习向量进行分析,表示学习向量维度的取值范围为 $\text{dim} = \{2000, 2500, 3000, 3500, 4000\}$,参数分析的结果如图 5(b) 所示。可以看出,综合 3 种不同比例训练样本划分下的表现,3000 维的知识图谱表示学习向量能够获得较好的模型性能。

基于上述参数分析,本文提出的知识图谱引导深度网络学习的遥感影像场景分类模型在后续试验中将 ω 固定为 0.0001,经过土地覆盖概念语义表达得到的遥感场景语义类别向量固定为 3000 维语义类别向量。

2.3 知识图谱引导深度网络学习机制有效性分析

为了进一步验证本文提出的知识引导学习模

块能够引导深度网络学习得到最佳的遥感影像场景分类性能,开展了消融试验来进行分析。

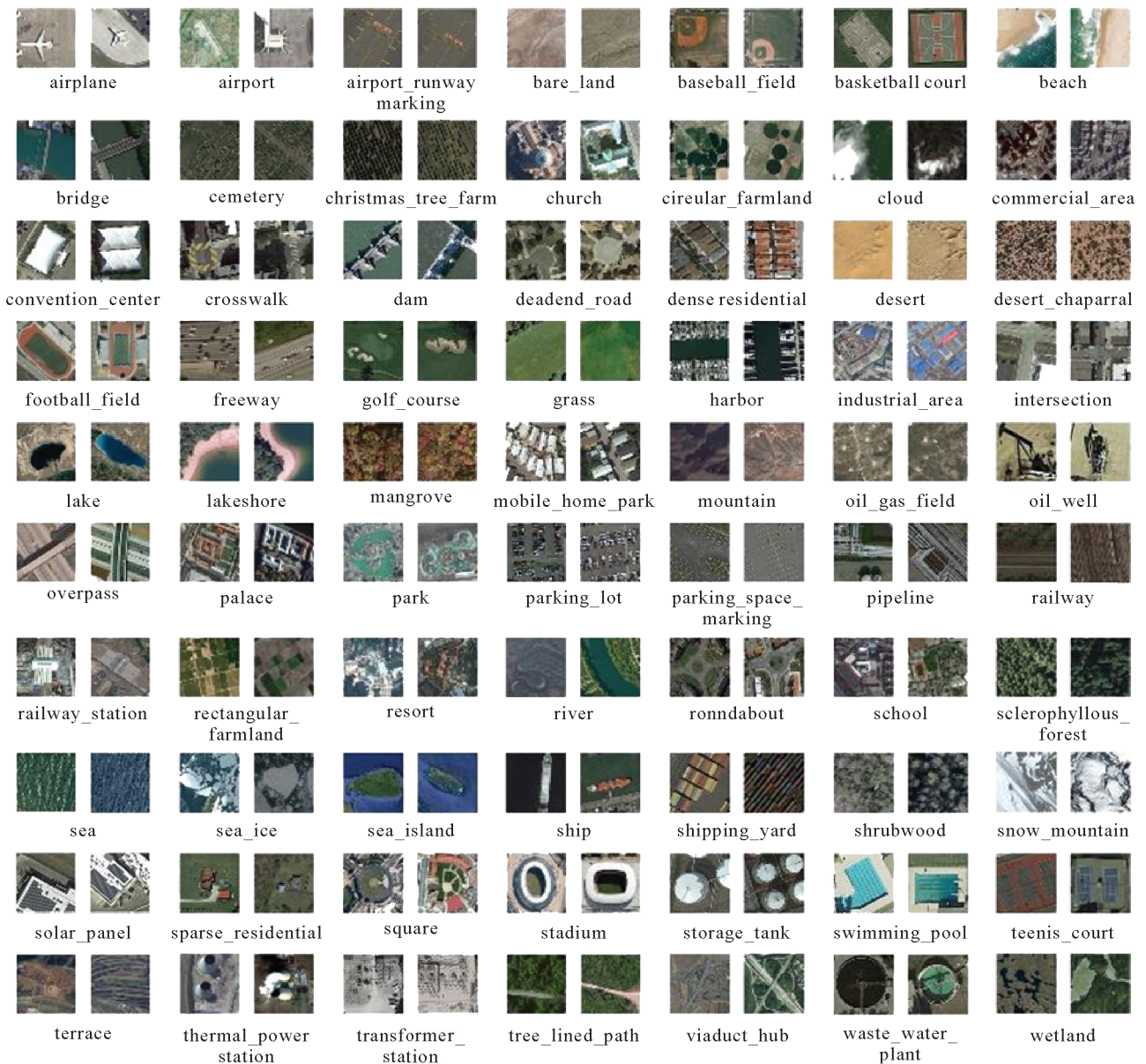


图 4 遥感影像场景数据集

Fig. 4 Remote sensing image scene dataset

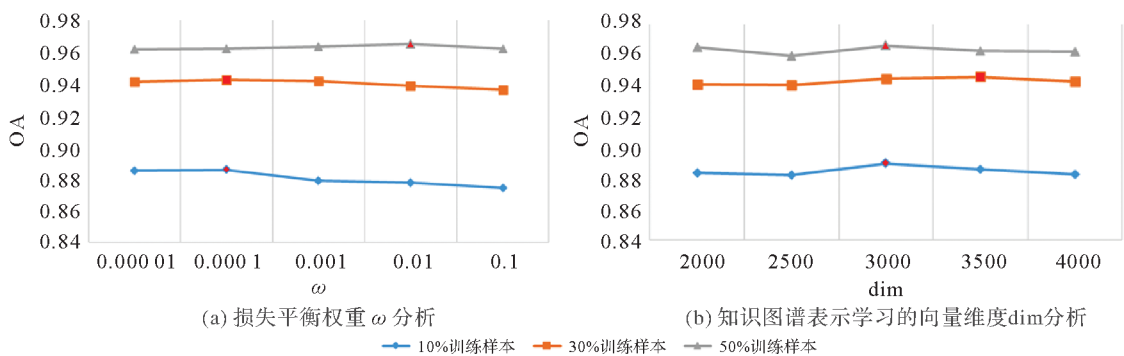


图 5 参数分析

Fig. 5 Parameter analysis

对于本文提出的知识引导学习模块的有效性,通过以下几种不同结构模型的试验来进行验证:①骨架网络模型 efficientnet-b3^[51] (EFNet);②对 EFNet 浅层部分通过“one-hot”场景类别标签进行精化的网络模型 EFNet-aux^[53];③对 EFNet 浅层部分通过 Word2Vec^[33] 语义向量进行知识引导学习的网络模型 (EFNet+Word2Vec);④对 EFNet 浅层部分通过 Bert^[53] 语义向量进行知识引导学习的网络模型 (EFNet+Bert);⑤对 EFNet 浅层部分通过 DistMult^[46] 表示学习得到的 3000 维语义向量进行知识引导学习的网络模型 KGE (EFNet+KGE);⑥本文提出通过知识引导学习模块和 DistMult^[46] 表示学习得到的 3000 维语义向量进行知识引导学习的网络模型。消融试验的定量结果如表 1 所示。可以看到,本文提出的土地覆盖概念语义表达得到的遥感场景语义基准相比于语义标签和两种自然语言处理模型得到的语义向量,均能够更好地引导深度网络进行学习。相较于表现相对最好的 Bert 模型语义向量,本文提出的遥感场景语义基准在 10%、30% 和 50% 训练条件下能够分别得到 0.75%、0.73% 与 0.21% 的精度提升。相较于骨架深度网络,本文提出的基于 VAE 的知识引导学习模块在 3 种不同训练条件下分别得到 5.11%、3.19% 与 2.22% 的精度提升,证明了本文方法引导深度网络学习的有效性。其中,在 10% 训练样本条件下能够得到最高的精度提升进一步说明本文提出的知识引导学习方法能够降低深度网络对训练样本的依赖。

表 1 不同分类方法下不同方法的准确率对比

Tab. 1 Comparison of accuracy of different methods under different classification methods (%)

消融试验模型	10%训练样本	30%训练样本	50%训练样本
EFNet	83.86	91.38	94.40
EFNet+One-hot	84.23(+0.46)	91.98(+0.60)	94.45(+0.05)
EFNet+Word2Vec	86.13(+2.27)	91.98(+0.60)	94.58(+0.18)
EFNet+Bert	86.87(+3.01)	92.09(+0.71)	94.64(+0.24)
EFNet+KGE	87.62(+3.76)	92.85(+1.47)	94.85(+0.45)
本文提出的知识引导学习方法	88.97(+5.11)	94.57(+3.19)	96.62(+2.22)

2.4 与已有方法的对比分析

为了验证本文方法的先进性,将本文提出的知识图谱引导深度网络学习方法与已有遥感影像场景分类方法进行了试验对比。对比方法包括基于卷积神经网络的遥感影像场景分类方法和基于视觉 Transformer 的遥感影像场景分类方法。

对于基于卷积神经网络的遥感影像场景分类方法,本文采纳的对比基线方法包括:EFNet-aux^[53] 在深度网络中间层增加场景类别标签约束,以提高深度网络浅层视觉特征提取的性能;SAFF^[15] 对卷积神经网络提取出的特征进行增强,提取出更多具有鉴别性的特征,增强模型在遥感影像场景分类任务中的有效性;EAM^[16] 通过在卷积神经网络的基础上增加一个增强注意力模块,使得模型能够通过注意力机制学习到更多具有鉴别力的特征;PCNet^[17] 通过对成对的遥感影像进行选取与特征比较,来捕捉影像特征间的微小差异,并基于这个微小差异获取更有区别力的图像特征。对于基于视觉 Transformer 的遥感影像场景分类方法,本文采纳的对比基线方法包括:v16_21k^[18] 将在 ImageNet 上预训练的视觉 Transformer 在遥感影像场景分类数据集上精炼,并进行遥感影像场景分类;SCViT^[19] 通过一个轻量级的通道注意力机制来弥补视觉 Transformer 仅考虑全局空间注意力的弊端。

表 2 给出了不同方法在本文数据集上的结果对比。可以看出,本文提出的知识图谱引导深度网络学习方法在各个比例的训练样本划分下都能取得最佳的精度表现。在训练样本相对不充分的 10% 训练样本条件下,EAM 这种基于注意力机制和 PCNet 这种通过影像对之间的对比学习的机制能够帮助深度网络学习到更有效的遥感影像视觉特征,缓解深度网络对训练样本的依赖,得到较好的分类精度。但是,本文通过先验知识能够引导深度网络学习到更有针对性的共享特征,在最终分类精度上能够优于上述方法。同时可以看到,本文将骨架深度网络替换为 PCNet 后,最终精度相对 PCNet 能够得到一定提升,证明了本文提出的知识引导学习模块具有较强的应用灵活性。

3 结论

本文着眼于现有遥感影像场景分类模型无法充分利用遥感场景先验知识的问题,通过构建包括领域标准、专家知识和通用知识的土地覆盖概念知识图谱,来更全面、更灵活、更便捷地管理与利用遥感场景先验知识。在土地覆盖概念知识图谱基础上,通过基于知识图谱表示学习的土地覆盖语义表达,形成遥感场景语义基准。通过遥感场景语义基准引导深度网络浅层部分学习更有效的共享特征,降低深度网络对训练样本的依赖。在现有遥感影像场景分类数据集上的试验表明,本文提出的遥感

场景语义基准相比于自然语言处理模型得到的语义向量具有更强的遥感场景先验知识表达能力,能够更好地引导深度网络进行共享特征学习。本文提出的知识引导学习模块能够有效提升深度网络的遥感影像场景分类能力,并且能显著降低深度网

络对于训练样本的依赖。与现有遥感影像场景分类方法对比,本文提出的知识引导学习方法展现出遥感影像场景分类性能上的先进性与应用上的灵活性。

表2 不同划分比率下不同方法的准确率对比

Tab. 2 Comparison of accuracy of different methods under different partition ratios

(%)

主干网络	遥感影像场景分类方法	10%训练样本	30%训练样本	50%训练样本
ViT	v16_21k ^[18]	81.90	83.87	88.92
	SCViT ^[19]	82.34	83.65	90.71
VGG16	SAFF ^[15]	83.81	89.42	91.21
	本文提出的知识引导学习方法(VGG16)	85.52	91.65	94.36
efficientnet-b3	EFNet-aux ^[54]	84.32	91.98	94.45
	本文提出的知识引导学习方法(EFNet)	88.97	94.57	96.62
ResNet-101	EAM ^[16]	87.37	92.76	93.26
	PCNet ^[17]	88.76	91.57	94.46
	本文提出的知识引导学习方法(PCNet)	90.53	94.48	96.09

参考文献

- [1] LI Yansheng, ZHANG Yongjun, ZHU Zhihui. Error-tolerant deep learning for remote sensing image scene classification[J]. IEEE Transactions on Cybernetics, 2021, 51(4): 1756-1768.
- [2] 龚希, 吴亮, 谢忠, 等. 融合全局和局部深度特征的高分辨率遥感影像场景分类方法[J]. 光学学报, 2019, 39(3): 0301002. GONG Xi, WU Liang, XIE Zhong, et al. Classification method of high-resolution remote sensing scenes based on fusion of global and local deep features[J]. Acta Optica Sinica, 2019, 39(3): 0301002.
- [3] 白坤, 慕晓冬, 陈雪冰, 等. 融合半监督学习的无监督遥感影像场景分类[J]. 测绘学报, 2022, 51(5): 691-702. DOI: 10.11947/J. AGCS. 2022. 20210270. BAI Kun, MU Xiaodong, CHEN Xuebing, et al. Unsupervised remote sensing image scene classification based on semi-supervised learning[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(5): 691-702. DOI: 10.11947/J. AGCS. 2022. 20210270.
- [4] ZHU Qiqi, ZHONG Yanfei, ZHANG Liangpei, et al. Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(10): 6180-6195.
- [5] HUANG Xin, HAN Xiaopeng, MA Song, et al. Monitoring ecosystem service change in the city of Shenzhen by the use of high-resolution remotely sensed imagery and deep learning[J]. Land Degradation & Development, 2019, 30(12): 1490-1501.
- [6] YAO Xiwen, HAN Junwei, CHENG Gong, et al. Semantic annotation of high-resolution satellite images via weakly supervised learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(6): 3660-3671.
- [7] LI Yansheng, CHEN Wei, HUANG Xin, et al. MFVNet: a deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation[J]. Science China Information Sciences, 2023, 66(4): 140305.
- [8] YANG Yi, NEWSAM S. Comparing SIFT descriptors and Gabor texture features for classification of remote sensed imagery[C]//Proceedings of the 15th IEEE International Conference on Image Processing. San Diego: IEEE, 2008: 1852-1855.
- [9] ZHONG Yanfei, ZHU Qiqi, ZHANG Liangpei. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(11): 6207-6222.
- [10] LI Yansheng, TAO Chao, TAN Yihua, et al. Unsupervised multilayer feature learning for satellite image scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(2): 157-161.
- [11] XU Kejie, DENG Peifang, HUANG Hong. Vision transformer: an excellent teacher for guiding small networks in remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 3152566.
- [12] XIA Guisong, HU Jingwen, HU Fan, et al. AID: a benchmark data set for performance evaluation of aerial scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965-3981.
- [13] CHENG Gong, HAN Junwei, LU Xiaoqiang. Remote sensing image scene classification: benchmark and state of the art[J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.
- [14] YANG Yi, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]//Proceedings of the 18th SIGSPA-

- TIAL International Conference on Advances in Geographic Information Systems. San Jose; ACM Press, 2010: 270-279.
- [15] CAO Ran, FANG Leyuan, LU Ting, et al. Self-attention-based deep feature fusion for remote sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(1): 43-47.
- [16] ZHAO Zhicheng, LI Jiaqi, LUO Ze, et al. Remote sensing image scene classification based on an enhanced attention module[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(11): 1926-1930.
- [17] ZHANG Yue, ZHENG Xiangtao, LU Xiaoqiang. Pairwise comparison network for remote-sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [18] BAZI Y, BASHMAL L, AL RAHHAL M M, et al. Vision transformers for remote sensing image classification[J]. Remote Sensing, 2021, 13(3): 516.
- [19] LÜ Pengyuan, WU Wenjun, ZHONG Yanfei, et al. SCViT: a spatial-channel feature preserving vision transformer for remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 3157671.
- [20] LI Lingjun, HAN Junwei, YAO Xiwen, et al. DLA-MatchNet for few-shot remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(9): 7844-7853.
- [21] LI Yansheng, KONG Deyu, ZHANG Yongjun, et al. Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 179: 145-158.
- [22] LI Yansheng, OUYANG Song, ZHANG Yongjun. Combining deep learning and ontology reasoning for remote sensing image semantic segmentation[J]. Knowledge-Based Systems, 2022, 243: 108469.
- [23] LI Yansheng, ZHOU Yuhan, ZHANG Yongjun, et al. DKDFN: domain knowledge-guided deep collaborative fusion network for multi-modal unitemporal remote sensing land cover classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 186: 170-189.
- [24] MA Xiaorui, WANG Hongyu, LIU Yi, et al. Knowledge guided classification of hyperspectral image based on hierarchical class tree[C]// Proceedings of 2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama; IEEE, 2019: 2702-2705.
- [25] ZHU Qiqi, LEI Yang, SUN Xiongli, et al. Knowledge-guided land pattern depiction for urban land use mapping; a case study of Chinese cities[J]. Remote Sensing of Environment, 2022, 272: 112916.
- [26] FENSEL D, SIMSEK U, ANGELE K, et al. Introduction: what is a knowledge graph? [M]// Knowledge Graphs. Cham: Springer, 2020: 1-10.
- [27] 田玲, 张谨川, 张晋豪, 等. 知识图谱综述: 表示、构建、推理与知识超图理论[J]. 计算机应用, 2021, 41(8): 2161-2186.
- TIAN Ling, ZHANG Jinchuan, ZHANG Jinhao, et al. Knowledge graph survey: representation, construction, reasoning and knowledge hypergraph theory[J]. Journal of Computer Applications, 2021, 41(8): 2161-2186.
- [28] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]// Proceedings of 2008 ACM SIGMOD international conference on Management of data. Vancouver; ACM Press, 2008: 1247-1250.
- [29] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia-a large-scale, multilingual knowledge base extracted from Wikipedia[J]. Semantic Web, 2015, 6(2): 167-195.
- [30] CHEN Jindong, WANG Ao, CHEN Jiangjie, et al. CN-Probase: a data-driven approach for large-scale Chinese taxonomy construction [C]// Proceedings of the 35th International Conference on Data Engineering (ICDE). Macao; IEEE, 2019: 1706-1709.
- [31] 李彦胜, 武康, 欧阳松, 等. 地学知识图谱引导的遥感影像语义分割[J]. 遥感学报, 2024, 28(2): 455-469. DOI: 10.11834/Jrs.20231110.
- LI Yansheng, WU Kang, OUYANG Song, et al. Geographic knowledge graph-guided deep semantic segmentation network for remote sensing imagery [J]. National Remote Sensing Bulletin, 2024, 28(2): 455-469. DOI: 10.11834/Jrs.20231110.
- [32] JANOWICZ K, HITZLER P, LI Wenwen, et al. Know, know where, KnowWhereGraph: a densely connected, cross-domain knowledge graph and geo-enrichment service stack for applications in environmental intelligence[J]. AI Magazine, 2022, 43(1): 30-39.
- [33] 张雪英, 张春菊, 吴明光, 等. 顾及时空特征的地理知识图谱构建方法[J]. 中国科学: 信息科学, 2020, 50(7): 1019-1032.
- ZHANG Xueying, ZHANG Chunju, WU Mingguang, et al. Spatiotemporal features based geographical knowledge graph construction [J]. Scientia Sinica (Informationis), 2020, 50(7): 1019-1032.
- [34] STADLER C, LEHMANN J, HÖFFNER K, et al. LinkedGeoData: a core for a web of spatial open data[J]. Semantic Web, 2012, 3(4): 333-354.
- [35] 李彦胜, 张永军. 耦合知识图谱和深度学习的新一代遥感影像解译范式[J]. 武汉大学学报(信息科学版), 2022, 47(8): 1176-1190.
- LI Yansheng, ZHANG Yongjun. A new paradigm of remote sensing image interpretation by coupling knowledge graph and deep learning[J]. Geomatics and Information Science of Wuhan University, 2022, 47(8): 1176-1190.
- [36] 张永军, 程鑫, 李彦胜, 等. 利用知识图谱的国土资源数据管理与检索研究[J]. 武汉大学学报(信息科学版), 2022, 47(8): 1165-1175.
- ZHANG Yongjun, CHENG Xin, LI Yansheng, et al. Research on land and resources management and retrieval using knowledge graph [J]. Geomatics and Information Science of Wuhan University, 2022, 47(8): 1165-1175.
- [37] 张永军, 王飞, 李彦胜, 等. 遥感知识图谱创建及其典型场景应用技术[J]. 遥感学报, 2023, 27(2): 249-266.
- ZHANG Yongjun, WANG Fei, LI Yansheng, et al. Remote sensing knowledge graph construction and its application in typical scenarios

- [J]. National Remote Sensing Bulletin, 2023, 27(2): 249-266.
- [38] MIKOLOV T, SUTSKEVER I, CHEN Kai, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe; ACM Press, 2013; 3111-3119.
- [39] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2024-01-05]. <http://arxiv.org/abs/1810.04805>.
- [40] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe; ACM Press, 2013; 2787-2795.
- [41] WANG Zhen, ZHANG Jianwen, FENG Jianlin, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec City; ACM Press, 2014; 1112-1119.
- [42] JI Guoliang, HE Shizhu, XU Liheng, et al. Knowledge graph embedding via dynamic mapping matrix[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Stroudsburg; IEEE, 2015; 687-696.
- [43] SUN Zhiqing, DENG Zhihong, NIE Jianyun, et al. RotatE: knowledge graph embedding by relational rotation in complex space. [EB/OL]. [2024-01-01]. <http://arxiv.org/abs/1902.10197>.
- [44] ZHANG Zhanqiu, CAI Jianyu, ZHANG Yongdong, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(3): 3065-3072.
- [45] NICKEL M, TRESP V, KRIEGEL H P. A three-way model for collective learning on multi-relational data[C]//Proceedings of 2011 International Conference on Machine Learning. New York; ACM Press, 2011.
- [46] YANG Bishan, YIH Wentau, HE Xiaodong, et al. Embedding entities and relations for learning and inference in knowledge bases[EB/OL]. [2024-01-01]. <http://arxiv.org/abs/1412.6575>.
- [47] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York; ACM Press, 2016; 2071-2080.
- [48] BALAZEVIC I, ALLEN C, HOSPEDALES T. TuckER: tensor factorization for knowledge graph completion[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg; IEEE, 2019; 5185-5194.
- [49] MARINO K, SALAKHUTDINOV R, GUPTA A. The more you know: using knowledge graphs for image classification[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE, 2017; 20-28.
- [50] SPEER R, CHIN J, HAVASI C. ConceptNet 5.5: an open multilingual graph of general knowledge[C]//Proceedings of 2017 AAAI Conference on Artificial Intelligence. San Francisco; ACM Press, 2017; 4444-4451.
- [51] TAN Mingxing, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks[C]//Proceedings of 2019 International Conference on Machine Learning. Long Beach; IEEE, 2019; 6105-6114.
- [52] LI Yansheng, ZHU Zhihui, YU Jingang, et al. Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(12): 10590-10603.
- [53] BAZI Y, AL RAHHAL M M, ALHICHRI H, et al. Simple yet effective fine-tuning of deep CNNs using an auxiliary classification loss for remote sensing scene classification[J]. Remote Sensing, 2019, 11(24): 2908.
- [54] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2024-01-01]. <http://arxiv.org/abs/1409.1556>.

(责任编辑:张艳玲)

Knowledge graph-guided deep network for high-resolution remote sensing image scene classification

LI Yansheng, WU Minlang, ZHANG Yongjun

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Abstract: Thanks to the rapid development of deep network theory and methods, deep networks have gradually become the mainstream technology for remote sensing image scene classification tasks. However, existing deep network-based remote sensing image scene classification methods are highly dependent on a large number of manually labeled training samples and cannot effectively integrate and utilize the rich prior knowledge in the remote sensing field. In order to improve the utilization of domain knowledge while reducing the dependence on labeled samples, this paper proposes a knowledge graph-guided deep network learning method for high-resolution remote sensing image scene classification. First, this paper constructs a land cover concept knowledge graph that includes various sources of knowledge in the field to more flexibly and conveniently apply domain prior knowledge. Furthermore, through the knowledge graph representation learning method, the semantic categories of remote sensing scenes in the land cover concept knowledge graph are expressed as semantic vectors to form a semantic benchmark for remote sensing scene categories. In the knowledge-guided learning stage, the cross-modal alignment constraint between the scene category semantic vector and the shallow visual feature vector of the deep network is applied to guide the shallow part of the deep network to more effectively learn shared features of different categories of remote sensing image scenes, while in the deep part of the deep network, it is still guided by scene category labels to learn discriminative features of different remote sensing scenes. In the testing stage, the optimized deep network model can complete high-precision remote sensing image scene classification without relying on any prior knowledge. The experimental results on the currently largest publicly available remote sensing image scene classification dataset show that the proposed knowledge-guided learning method can obtain optimal classification performance at different training sample ratios such as 10%, 30%, and 50% compared with existing methods. Under the condition of 10% sample ratio, our proposed method can achieve an improvement of 5.11% in overall accuracy (OA) compared with baseline deep networks.

Key words: remote sensing image scene classification; land cover concept knowledge graph; knowledge graph representation learning; knowledge graph-guided deep network

Foundation support: The National Natural Science Foundation of China (Nos. 42030102; 42371321)

First author: LI Yansheng (1987—), male, PhD, professor, majors in remote sensing spatio-temporal knowledge graph, multi-modal remote sensing foundation model, and intelligent interpretation of remote sensing big data.

E-mail: yansheng.li@whu.edu.cn